Psychometric Analysis of the Low Vision Diagnostic Assessment Procedure

Edward P. Berla', Ph.D.
University of Louisville
American Printing House for the Blind

Earl F. Rankin, Ph.D.
University of Kentucky

Debbie H. Willis, B.A.
American Printing House for the Blind

Running Head: Psychometric Analysis

Abstract

A psychometric analysis was conducted on the Diagnostic Assessment Procedure (DAP) which evaluates the visual efficiency of low vision legally blind persons. The DAP was administered twice to each of 112 legally blind students ages 5-20. Reliability coefficients indicate that the DAP was highly internally consistent (KR-20 = .94) and highly accurate (test-retest, $\underline{r}$ = .96). The hierarchical nature of the DAP and its content and construct validity were evaluated and are discussed.

Diagnostic Assessment Procedure

A Program to Develop Efficiency in Visual Functioning is currently being developed by the American Printing House for the Blind in collaboration with Barraga and her colleagues (Barraga & Collins, 1979; Barraga, Collins, & Hollis, 1977). This program grew out of Barraga's earlier work (Barraga, 1964) which demonstrated that legally blind persons with severe visual handicaps could be systematically taught to increase the efficiency with which they use their remaining vision. Subsequent research by Ashcroft, Halliday, & Barraga (1965) and Nolan (1967) replicated and extended Barraga's work. The program currently under development represents a substantive modification of the earlier work.

The Program currently under development consists of a 40 item assessment procedure, a set of 150 lessons which are keyed to specific sections of the assessment procedure, and accompanying source book. The purpose of the assessment procedure is to evaluate a person's current level of visual functioning in order to plan a systematic program of visual training designed to increase the efficiency with which the person will use his/her remaining vision. The assessment procedure can also be used to determine progress following a period of training.

The purpose of this article is to report the outcome of a psychometric analysis of the low vision assessment procedure which was carried out during the fall of 1978.

Subjects

The subjects for this study were selected on the basis of the following criteria:

1. a minimum developmental level of 5.5 years in age
2. a minimum of 1 year in a structured educational setting (kindergarten and up)
3. legally blind (20/200 visual acuity or less in the better eye with corrective lens or a visual field of 20 degrees or less) with a minimum of light perception
4. no additional physical or mental handicaps that would interfere with performance on the test

The subjects were 112 legally blind pupils attending schools in the United States (California, Georgia, Kansas, New Hampshire, New York, Oregon, Texas, Massachusetts, Washington) and Canada (Ontario). All subjects were volunteers. There were 60 males (54%) and 52 females (46%) with 89 subjects from public schools (79%) and 23 subjects from residential schools (21%). Each subject was classified into one of nine visual categories using the classification system of visual acuity proposed by the National Society for the Prevention of Blindness. For the purposes of this study another category (X) was added to take into account those subjects who had only restrictions in visual fields of less than 20° but whose visual acuities were greater than 20/200. The number and percentage of subjects in each of the visual categories is shown in Table 1. These proportions compare favorably to those reported

Insert Table 1 about here

by Willis (1977) using the 1976 quota registration figures of the legally blind students attending residential and public schools in the United States as compiled by the American Printing House for the Blind. Consequently, the results of this study can be generalized to the legally blind population of school age students. The reader should note that totally blind children were excluded from this study and consequently, there were no students classified in category IX.

The incidence of the eye disorders/causes represented in the subjects is shown in Table 2. The reader should note that the total incidence of 217 exceeds the sample size of 112 because most of the subjects were diagnosed as having more than a single eye disorder. The large number of eye disorders in the sample attests to the great heterogeneity of visual disabilities in the population.

The subjects' ages ranged from 5-20 years. The median age was [illegible] years.

Insert Table 2 about here

Method

Materials

Diagnostic Assessment Procedure (DAP)

The DAP consists of 40 items ranging from the simple item of turning a light to the most difficult items of writing letters and reading

simple words. The 40 items of the VP are broken down into eight categories that correspond to specific functional visual tasks. Because certain categories were more heterogeneous than others, a different number of questions were required to adequately sample the behaviors in the different categories. The eight categories with the number of items in each category (shown in parentheses), along with examples of test objectives, are as follows:

A. Awareness of visual stimuli (2)

Item No. 1. "Learner moves eyes, head and/or body to light source."

B. Movement control of eyes, discrimination of shape and color (7)

Item No. 3. "Learner will look from one light to the other."

C. Exploration, discrimination, use of objects (6)

Item No. 7. "Learner will see two lines and move between the

D. Discrimination and identification of pictures of objects, people, actions (12)

Item No. 14. "Learner will match solid color geometric shape in pictures."

E. Memory for detail, part whole relationships, figure-ground discrimination (4)

Item No. 25. "Learner will identify object partially hidden in picture."

F. Discrimination, identification, reproduction of abstract figures and symbols (4)

Item No. 29. "Learner will match abstract figures by single inner detail."

G. Perception [illegible] bols (4)

Item No. 33. "Learner will relate inner detail in figures of different size."

H. Identification, perception, reproduction of symbols - single and combined symbols (4)

Item No. 37. "Learner will identify [illegible]

Each item specifies the visual task, the materials, the [illegible] for administering the item, and the specific response required of the learner in order to be considered correct.

All materials necessary for administering an item are self-contained.

Procedure

Twelve teachers of the visually handicapped were selected to be evaluators. The criteria for selection were as follows.

1. Interest in participating
2. Experienced in testing procedures, specifically in using the Visual Efficiency Scale (Barraga, 1965) for the assessment of low vision
3. Access to at least 10 students who met the criteria for participation in the study.

Each teacher was required to attend a 2 1/2 day training session held at the American Printing House for the Blind during the summer of 1972. During the training session the rationale of the program was given along with a demonstration in administering the entire assessment procedure.

The teacher-evaluators returned to their respective centers to assess their students. Each teacher-evaluator was asked to test and retest each student within a period of not less than 2 nor more than 4 weeks during the fall of 1978. The teacher-evaluators were paid for their participation in the training session and for each set of test-retest protocols that were returned to APH.

Results and Discussion

This instrument was examined from both a content validity as well as from a psychometric standpoint. Both these standpoints are essential to the users of this instrument because a person's performance on the categories of the test are more important to the examiner than the total test score (i.e., educational planning and educational intervention will be based on performance within the eight categories rather than on the test as a whole).

The content validity of the Diagnostic Assessment Procedure is quite good. In constructing the test the developers reported reviewing over 100 articles and books on visual perception, normal visual development, visual impairment, and low vision. In addition, 14 professionals in the areas of visual perception, low vision, and teachers of the visually handicapped served as consultants on the project. The literature review and the input from the consultants provided the basis for the development of the eight functional visual categories and the specific items of the instrument. The eight categories of the instrument appear to be very comprehensive in covering the full range of visual functioning. In addition

the items of the [illegible] representative of [illegible] that would be necessary in performing a wide range of actual real-life visual tasks.

The means, standard deviations, and standard errors of [illegible] for the test-retest protocols are shown in Table 3. The average subject obtained a score of approximately 70% on both the test and retest indicating that the DAP was rather easy. A statistical test [illegible] ness of the distribution was significant ($p < .05$). Visual inspection of the distribution showed a negative skewed curve with most scores piled up at the high end of the distribution.

Insert Table 3 about here

A comparison of the means and standard deviations for both the test and retest demonstrated a high degree of consistency of visual functioning over the period between test and retest with respect to both the level and degree of homogeneity among individuals in the group. The small standard error of the mean indicates that there is a probability of $p < .01$ that the mean score for the population is less than two points on either side of the test or retest means.

Item Analysis

No negatively discriminating items were found as a result of either the test or retest. However, there were a few items which deviated substantially in difficulty from other items on the test which warranted special attention. Table 4 shows the list of difficult items for both the test and retest. It should be noted that the same items were [illegible]

to be difficult on both the test and retest. An analysis of these items revealed that they tended to demand drawing and graphic manipulative skills which predictably would be difficult for persons with visual handicaps.

Insert Table 4 about here

Reliability estimates for the DAP were based upon internal consistency coefficients (i.e., Kuder-Richardsen 20) and correlations between test and retest results. The reliability coefficients and standard errors of measurement are shown in Table 5. The internal consistency of both the test and retest were identical and very high, $r = .94$. The errors of measurement were extremely small, and it can be concluded that an individual's true visual efficiency score would probably ($p < .05$) be found within plus or minus two points on either side of his score as measured by this test. The test-retest correlation was $r = .96$ which indicates that the DAP is a highly accurate instrument and that the type of visual functioning measured by the DAP does not vary much from week to week.

Insert Table 5 about here

Hierarchial Sequence

It could be hypothesized that a hierarchical sequence from easy to difficult items exists in the assessment procedure. Such a hierarchy may exist in one or both of two types. One type of hierarchy (Type I

from the beginning to the end of the instrument. This type of hierarchy would produce a progressive decrement in the mean percentage [illegible] for categories I through M which would consist of clusters of items organized sequentially from the beginning to the end of the instrument. A second type of hierarchy (Type B) would be based on the assumption that a degree of interdependency exists among items (or categories). Success at lower levels in the structure would be necessary before an individual could respond adequately to items at higher levels in the hierarchical structure. A Type A hierarchy is a necessary but not a sufficient condition for a Type B hierarchy.

Evidence for a Type A hierarchy

1. Hierarchy of item difficulty

Space precludes the description of item difficulties for all 40 items, but a visual inspection of item difficulty percentages in the item analysis print-out sheet revealed a slight but nonsignificant tendency (Kruskal-Wallis test, $p > .05$) for later items to be slightly more difficult than early items in the sequence. However, a number of erratic fluctuations in difficulty are in evidence from the beginning to the end of the instrument. Consequently, this evidence does not support the notion of a hierarchy of item difficulty

2. Hierarchy of category difficulty

If there is a hierarchy of category difficulty then there should be increasing difficulty across categories. Table 6

presents mean percentage scores for category A on the initial test and the retest results.

Insert Table 6 about here

The rank order of mean percentages for the initial test ranges in a descending manner from a high of 94% for category A to a low of 45% for category H and thus provides evidence for the existence of a Type A hierarchy for the category structure. The retest results are similarly hierarchical in structure with the exception of category E. However, the initial test provides better evidence for the existence of a hierarchy because it was not influenced by potential practice effects that may have occurred on the retest.

Evidence for a Type B Hierarchy

If a state of dependency exists among items (or categories) in a hierarchial structure such that success on tasks at a given level in the hierarchy is dependent upon the mastery of other tasks at lower levels, then correlations between items at a particular level and items at successively higher levels should become progressively lower. Similarly, correlations between items at a given level and other items at successively lower levels should become progressively lower. Two different types of analyses were conducted as follows:

1. Items--Tetrachoric correlations were computed among all possible combinations of items. Such correlations describe the strength of the relationship between passing versus failing on one item and passing versus failing on another item. Space does not

permit [illegible] to [illegible] correlations. [illegible] can be reported that no trend existed among these correlations in keeping with the assumptions of a Type B hierarchy among items in the test.

2. Categories--Tables 7 and 8 provide evidence related to the possibility of a Type B hierarchy among categories A-H.

Insert Table 7 about here

The inter-correlations shown in Table 7 show a general tendency for correlation trends to conform to predictions made from the assumption of a Type B hierarchy. The encircled correlation coefficients in Table 7 represent deviations from the predicted trends. A summary of results in Table 8 shows that 73% of the correlations were in keeping with the hierarchial hypothesis. However, the correlations among all categories (including those far removed from a given category level) are too high to permit the conclusion that an individual who does not succeed at a particular level cannot succeed at higher levels of visual functioning.

Insert Table 8 about here

In summary, for individual items, no evidence exists to support the assumption of either type of hierarchy. At the level of categories, however, it can be concluded that a Type A hierarchy exists but not a Type B hierarchy. In other words, significant differences in difficulty exist as one proceeds from the early category of items to the later ones.

in this test. However, [illegible] exceptions exist to this rule [illegible] categories cannot be responded to adequately by an individual who does not succeed at lower categories.

Construct Validity of Categories A-H

A factor analysis was conducted on the initial test results. The factor analysis revealed that one factor accounted for 91% of the variance and a second factor accounted for about 5% of the variance. Three other factors were indicated, but were probably not valid. Most of the items were loaded on factor I. The second factor had high factor loadings on 17 items. Without going into a detailed description of these results, it can be said that there was no correspondence between the factor loading for individual items and the hypothesized categories A-H. If these categories had construct validity, then items for a given category would load highly on that category but would show low loadings on other categories. Such results were not found.

Visual inspection of inter-correlations among categories reveals modest and significant correlations among all categories. In fact, the correlations between test and retest results for the same categories were not substantially larger than the correlations among different categories for the same test.

It can therefore be concluded that these categories lack construct validity. There is little evidence that this test measures any clusters of traits corresponding to the hypothesized category system. From a measurement point of view, even if these categories were valid, due to the high correlations among all categories, the error of measurement

among categories would be too large to permit the diagnostic interpretation to be placed upon differences among category scores that such differences were rather large.

It should be said, however, that the requirements for [illegible] construct validity are so rigorous that few, if any, psychological or educational measuring instruments meet such criteria today. Most published reading tests, for example, do not demonstrate sufficient subtest scores to be used for diagnostic purposes. Factor analytic studies of such tests seldom reveal factors corresponding to most subtest titles. Yet, these tests are obviously useful diagnostic tools. They may display adequate content validity and criterion related validity despite the lack of factorial validity.

Paradoxically, despite the previously discussed results concerning the lack of construct validity of this instrument, it would nevertheless be feasible to use this category system as an organizational tool for relating assessment results to treatment procedures. The rationale for this decision would rest upon both the substantial content validity of this instrument and an expected Type A hierarchy among categories. The content of this instrument was based upon a rational analysis by professionals in the field of the visually handicapped as well as psychologists who were knowledgeable in the field of visual perception.

If the category structure were to be used as a diagnostic tool justified on pragmatic if not entirely on rigorous psychometric grounds, it would be necessary that these categories possess a reasonably high degree of reliability and a low degree of measurement error. Reliability

in Tables 6 and 9 provide [illegible] decision.

Insert Table 9 about here

It can be seen that the reliability coefficients in Table 9 are surprisingly high (considering the small number of items in each category) and the errors of measurement are very low. Also, the [illegible] between test and retest means in Table 6 indicates a remarkable stability in test performance within each category. It should be noted, however, that the reliability coefficients (see Table 9) for categories F and G are lower than the coefficients for the other categories. Careful study should be given to the items in these two categories. Results for categories F and G shown in Tables 7 and 9 also suggest that an unidentified problem exists among the items in these categories. That is to say, the predicted correlation trends in these Tables for categories F and G are less precise than for most other categories.

One of the limitations of the preceding analysis concerns the nature of the population sample. The results were limited to those individuals who were selected on the basis of visual acuity of 20/200 or less with no additional handicaps (mental or physical) that might interfere with their performance on the test. Furthermore, since the population of visually handicapped persons is quite heterogeneous with respect to the nature of the eye disorder as well as visual acuity and visual field, it is an open question as to the nature of the relationship between these visual characteristics and performance on the

test. Empirically, this is a difficult problem [illegible] the small number of persons that would fall into any category.

In conclusion, this is a highly accurate instrument which displays a high level of content validity for its purpose. More evidence is needed concerning the correlation of performance on the instrument with related variables to determine its criterion related [illegible] Within the context of this program the Diagnostic Assessment Test instrument seems to be well designed for making effective use of learning procedures for individuals.

Ashcroft, S. C., Halliday, C., & Barraga, N. Study II: Effects of experimental teaching on the visual behavior of children evaluated as though they have no vision. Nashville, Tenn.: George Peabody College for Teachers, 1965 (USOE Grant No. 32-52-1021-1030).

Barraga, N. C. Increased visual behavior in low vision children. American Foundation for the Blind Research Series, 1964, No. 13.

Barraga, N. C., Collins, M., & Hollis, V. Development of efficiency in visual functioning: A literature analysis. Journal of Visual Impairment & Blindness, 1977, 71, 337-391.

Barraga, N. C., & Collins, M. E. Development of efficiency in visual functioning: Rationale for a comprehensive program. Journal of Vision Impairment & Blindness, 1979, 73, 121-126.

Holmes, R. V. Training residual vision in adolescents educated previously as nonvisual. Unpublished master's thesis, Illinois State University, 1967.

The research reported herein was performed pursuant to a grant (Grant No. G007[illegible]48, Revision of the Utilization of Low Vision Kit) with the Bureau of Education for the Handicapped, U.S. Office of Education, Department of Health, Education and Welfare to the American Printing House for the Blind, Louisville, Kentucky. The authors express their appreciation to June Morris, Director of the Department of Educational Research, and Amie Dennison, Research Associate, both of the American Printing House for the Blind and Natalie Barraga from the University of Texas for their aid throughout the entire project. Opinions expressed in this article are those of the authors and do not represent the official position of the Bureau of Education for the Handicapped.

Table 1

The Number and Proportion of Subjects in Each of the Visual Categories and the Proportion of Subjects in the 1976 Quota Registration Reported by the American Printing House

		Number	Proportion in sample	Proportion of legally blind (1976 quota)
I	20/200 - 10/200	59	.54	.[illegible]
II	70/200 - 30/200	5	.05	.[illegible]
III	40/200 - 8/200	21	.19	.[illegible]
IV	7/200 - 3/200	4	.04	.[illegible]
V	2.5/200 - .4/200	1	.04	.[illegible]
VI	Counts finger	7	.06	.[illegible]
VII	Hand movements (includes "forms" and object perception)	5	.05	.[illegible]
VIII	Light projection and/or perception (includes shadows)	6	.05	.12
IX	Totally blind	0	-	.18
X	Restricted visual field (20 degrees or less)	2	.02	
		110		

Note. The information on two subjects' visual acuity was [illegible] and consequently the data reported in this table were based on 110 subjects instead of 112. The reader should note that the visual categories in this study were slightly different from those used by Willis (1979). If the Willis data were recalculated to conform to our category system, it would slightly increase all of the percentages she reported. As a consequence, our sample proportions would approximate her population proportions even more closely than they already are.

Table I

Diagnoses and/or Causes of the Eye Disorders Represented in the Testing Sample (Arranged by Incidence)

	Name Given	Incidence
1.	Nystagmus	38
2.	Cataracts	22
3.	Myopia	15
4.	Optic Atrophy	13
5.	Albinism	11
6.	Aniridia	9
7.	Aphakia (5 indicated as surgical)	6
8.	Glaucoma	8
9.	Coloboma	7
10.	Esotropia	6
11.	Microphthalmos	6
12.	Retrolental Fibroplasia (RLF)	5
13.	Amblyopia	4
14.	Enucleation	4
15.	Retinitis Pigmentosa	4
16.	Rubella	4
17.	Exotropia	3
18.	Myopic Astigmatism	3
19.	Retinoblastoma	3
20.	Aebiomatopsia	2
21.	Astigmatism (see 30, also 18)	2

	Name Given	Incidence
22.	Corneal Scarring	2
23.	Myopropia	2
24.	Hypoplasia	2
25.	Macular Degeneration	2
26.	Optic Nerve Damage	2
27.	Retinitis	2
28.	Abnormal Macula	1
29.	Abnormal Optic Nerve Function	1
30.	Anomalies (multiple)	1
31.	Anterior Cleavage Syndrome	1
32.	Batton's Disease	1
33.	Choroiditoid	1
34.	Coats' Disease	1
35.	Corneal Vascularization	1
36.	Gun Wound	1
37.	Histoplasmosis	1
38.	Hyperopic Astigmatism	1
39.	Hyperplastic Primary Vitreous	1
40.	Microcornea	1
41.	Myopic Degeneration	1
42.	No Diagnosis	1
43.	Photophobia	1
44.	Prosthesis (same as 14)	

36.	Ptosis	1
37.	Retinal degeneration	1
47.	Retinal dysplasia Schisis of Retina (see 27)	1
48.	Strabismus	1
49.	Tapeto	1
50.	Toxoplasmosis	1
51.	Tumor	1
52.	Viral Encephalitis	1
	Total	211

Table 2

Descriptive Statistics for the Test and Retest Assessment

Assessment	Mean	Standard deviation	Standard error of mean
Test	28.4 (71%)	8.5	.68
Retest	28.9 (72%)	8.1	.75

Table 4

Difficult Items in Test and Retest Assessments

Test assessment		Retest assessment	
Item	Difficulty (%)	Item	Difficulty (%)
9	25	9	17
17	22	17	17
30	27	30	9
34	12	34	11
37	26	37	22
38	41	38	31

Psychometric Re

Table 2

Reliability Coefficients and Standard Errors of Measurement for the Test and Retest Assessment

Assessment	KR 20	Test-retest correlation
Test	.94	.86
Standard error	2.7	3.7
Retest	.94	
Standard error	1.9	1.6

Table 1

Mean Percentages for Categories A-H

on the Initial Test and Retest

Category	Initial test mean	Retest mean	Number of items	Initial test mean percentage	Retest mean percentage
A	1.87	1.91	2	94	96
B	3.59	3.70	4	90	93
C	4.64	4.63	6	77	77
D	8.88	9.25	12	74	77
E	2.73	2.94	4	69	73
F	2.69	2.50	4	67	62
G	2.24	2.33	4	56	58
H	1.79	1.73	4	45	43

N = 112

Table 1

Correlations Among Categories A-H

on the Initial Test

Categories	A	B	C	D	E	F	G
	X						
A	1.00	.72	.67	.59	.43	(.61)	.44
	←	X					
B	.72	1.00	.71	(.76)	.60	(.70)	.60
	←		X				
C	.67	.71	1.00	.72	.61	(.64)	.54
	←			X			
D	.59	(.76)	.72	1.00	.76	.75	.72
	←				X		
E	.43	.60	.61	.76	1.00	.62	(.68)
	←					X	
F	.61	(.70)	(.64)	(.75)	.62	1.00	.64
	←						X
G	.44	(.60)	.54	(.72)	(.68)	.64	1.00
	←						
H	.25	.38	.40	(.58)	(.53)	.52	.63

N = 112

←→ Arrows indicate direction of predicted trend.

( ) Circles indicate deviation from predicted trends.

Table 1

Percentage of Correctly Identified Words

on the Initial Test

Category	Percentage
A	85
B	71
C	86
D	85
E	71
F	57
G	57
H	71

Reliability Coefficients and Standard Errors of Measurement for Categories A-H

Category	Reliability*	SD	Standard error of measurement
A	.83	.41	.17
B	.84	.91	.36
C	.80	1.31	.59
D	.89	2.84	.94
E	.86	1.26	.62
F	.76	.97	.47
G	.69	1.13	.55
H	.83	1.44	.59

N = 112

* Correlations between pre and post assessment.

** Based on pre assessment results.